

Analysis

Mendelian randomization analysis reveals the potential of the IKZF1 gene as a therapeutic target in colorectal cancer

Junzuo Lin¹ · Qian Xu¹ · Zhengfei Zhao²

Received: 15 November 2024 / Accepted: 12 May 2025

Published online: 21 May 2025

© The Author(s) 2025 **OPEN**

Abstract

Introduction Colorectal cancer (CRC) is a global health burden, highlighting the urgent need for the discovery of new biomarkers and therapeutic targets. This study integrates genetic epidemiology methods, such as Mendelian randomization (MR), with GWAS data to predict treatment efficacy and identify novel CRC therapeutic targets.

Methods We utilized cis-eQTL data from the eQTLGen consortium and CRC GWAS data from the IEU Open GWAS database. MR analysis was conducted via the R package TwoSampleMR. Bayesian colocalization analysis was applied to identify shared genetic effects between CRC risk factors and potential therapeutic targets. Phenome-wide association study (PheWAS), protein–protein interaction (PPI) network construction, and enrichment analyses were performed to elucidate the functional profiles of the targets. Molecular docking and dynamics simulations were employed to evaluate the therapeutic potential of the identified targets.

Results MR analysis identified 60 genes associated with CRC risk. Our analysis identified IKZF1 as a significant therapeutic target through colocalization analysis. The PheWAS results revealed no significant genomic correlations for IKZF1, suggesting its potential as a specific therapeutic target. PPI and enrichment analyses highlighted the role of IKZF1 in epigenetic regulation and transcriptional control. Molecular docking and dynamics simulations confirmed the strong binding affinities of potential drugs with IKZF1.

Conclusion This study identified IKZF1 as a promising therapeutic target for CRC through MR and colocalization analyses. The target's association with immune modulation and epigenetic mechanisms, supported by molecular docking and dynamics simulations, positions IKZF1 as a key player in advancing precision CRC therapies, warranting further clinical investigation.

Keywords Colorectal cancer · Mendelian randomization · Gene colocalization analysis · Molecular docking · Drug target discovery · Phenotype-wide association study · Molecular dynamics simulations

Abbreviations

CRC	Colorectal cancer
GWAS	Genome-wide association studies
SNP	Single-nucleotide polymorphism
MR	Mendelian randomization

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12672-025-02683-0>.

✉ Zhengfei Zhao, yaohui19742022@163.com; Junzuo Lin, 18190035409@163.com; Qian Xu, 1121019227@qq.com | ¹Department of Graduate School, Southwest Medical University, Luzhou 64600, China. ²Department of Gastrointestinal Surgery, The Affiliated Hospital of Southwest Medical University, Luzhou 64600, China.



eQTL	Expression quantitative trait loci
cis-eQTLs	Cis-expression quantitative trait loci
PheWAS	Phenome-wide Association Study
IVW	Inverse variance weighted
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
PPI	Protein–protein interaction

1 Introduction

Colorectal cancer (CRC) is the third most prevalent cancer and a leading cause of cancer-related death globally [1]. It arises from factors such as obesity, tobacco use, and genetics [2]. Despite treatments, including surgery, chemotherapy, and immunotherapy, challenges such as postoperative relapse, drug resistance, and adverse effects remain. Thus, discovering new biomarkers and therapeutic targets is vital for enhancing early diagnosis and survival.

Adding genetic knowledge to drug development may increase innovation success, as personalized treatments often yield better clinical results [3–5]. Gene products have become drugs or drug candidates, including small molecules or antibodies [6, 7]. While GWASs (genome-wide association studies) have shown considerable effectiveness in pinpointing single nucleotide polymorphisms (SNPs) linked to the risk of CRC [8–10], they sometimes fall short in precisely pinpointing pathogenic genes that could be instrumental in the drug development process.

Mendelian randomization (MR), a method grounded in genetic epidemiology, emulates the structure of randomized controlled trials to predict the potential therapeutic efficacy of pharmaceuticals [11–14]. SNPs, recognized as expression quantitative trait loci (eQTLs), which are linked to alterations in gene expression, could mirror the enduring impacts of medications on their corresponding proteins [6, 15]. Furthermore, GWAS data pertaining to CRC can be harnessed to pinpoint distinct genetic variations (SNPs) correlated with this condition. By leveraging MR, researchers can amalgamate findings on the relationships between SNPs and gene expression with those between SNPs and disease, thereby establishing causal pathways between risk factors and diseases. Furthermore, after fertilization, individuals are randomly assigned genes that are targets for drugs with high or low expression levels, and individuals are typically unaware of their genotypes; therefore, MR studies are analogous to blinded experiments [16]. However, MR has limitations, such as potential confounding from cross-sectional effects, which can distort causality estimates. Consequently, these constraints and possible biases should be meticulously evaluated when assessing the findings from MR studies.

In our study, we identified novel CRC therapeutic targets via MR and GWAS data, establishing a causal pathway via cis-eQTLs to predict treatment efficacy. Furthermore, we conducted a gene colocalization analysis to substantiate the shared effects between potential therapeutic targets and CRC risk. This approach facilitated the elucidation of the causal nexus between treatment targets and the disease while mitigating the impact of confounding variables. Moreover, by employing a phenome-wide association study (PheWAS), we investigated the connections between potential therapeutic targets and a spectrum of other phenotypes, offering critical perspectives for the advancement of therapeutic strategies and highlighting their pleiotropic effects and potential therapeutic mechanisms. Concurrently, analyses of protein–protein interaction networks and gene set enrichment revealed the functional profiles and biological importance of prospective therapeutic targets, thereby deepening our understanding of the role of IKZF1 and its underlying mechanisms in the progression of colorectal cancer. Moreover, we employed pharmacophore modeling and molecular docking simulations to determine the therapeutic potential of IKZF1 within the context of CRC. We conducted a thorough feasibility study on the target and identified promising drug candidates by assessing their binding interactions and pharmacological effects. The blind docking technique is capable of forecasting potential protein binding sites, even those unoccupied in existing experimental models, and can project the most favorable positioning of ligands within these binding pockets, a key aspect in deciphering the intricacies of ligand–protein interactions [17, 18]. Molecular dynamics is a computational tool for modeling the motion of molecular systems. It generates trajectories of molecules over time by calculating the interaction forces between atoms and the equations of motion, thereby sampling configuration space. Its applications include conformational analysis, protein homology modeling, elucidation of biochemical and biophysical mechanisms, docking of biomolecular complexes and calculation of free energy changes, tasks that are crucial in structure-based drug design.

In summary, our research provides pivotal knowledge for pinpointing innovative treatment targets in colorectal cancer. Through the integration of these approaches, we have furnished a valuable framework for the formulation of more targeted and efficacious treatment approaches.

2 Materials and methods

2.1 Exposure data

Data on gene eQTLs were obtained from the eQTLGen consortium (<https://eqtlgen.org/>). Briefly, the consortium covers 16,987 genes and 31,684 eQTLs associated with cis-acting eQTLs in blood samples from healthy European populations. Specific information on these data can be found in Ref. [19]. In this research, we utilized the cis-eQTL dataset and allele frequency data from the consortium. Our drug target genes were identified by integrating GWAS data through computational methods from previous studies to find potential targets and their associations with drugs, resulting in the identification of 5431 candidate genes [7].

2.2 Outcome data

CRC GWAS data (ieu-b-4965) were sourced from the Integrative Epidemiology Unit (IEU) Open GWAS database (<https://gwas.mrcieu.ac.uk/>), encompassing 11,738,639 SNPs from 5,657 CRC patients and 372,016 controls of European descent, sourced from the UK Biobank.

2.3 Mendelian randomization analysis

The application of the R package TwoSampleMR (version 0.5.10) facilitated the MR analysis [20]. The R environment's `harmonize_data` function was used for dataset collection and normalization in our MR analysis. Genetic instrumental variables (IVs) for MR analysis are required to adhere to three MR assumptions: (1) a direct association between the SNP and the exposure (i.e., a significant correlation with gene expression, with a false discovery rate (FDR) below 0.05); (2) no association between the SNP and potential confounders of the exposure–outcome relationship; and (3) the SNP's influence on the outcome is mediated solely through the exposure [21]. This study applied quality control by limiting datasets to Europeans to reduce stratification bias and excluding IVs with F statistics ($F = (\beta/SE)^2$) below 10 to avoid weak IV bias [22].

MR studies are typically conducted via six distinct methodologies: MR–Egger, inverse variance weighting (IVW), the weighted median, simple mode, the Wald ratio and the weighted mode. A *P* value of less than 0.05 is typically regarded as indicative of statistical significance in MR analyses. An odds ratio (OR) > 1 indicates a risk factor, whereas an OR < 1 signifies a protective factor. In the primary analysis, MR estimates for single SNPs are computed via the Wald ratio method; for genes with multiple genetic instruments, SNP estimates are meta-analyzed via IVW, MR–Egger, and weighted median approaches. IVW assumes each genetic instrument's validity, offering the strongest statistical power [23]. The stability of the MR study outcomes was examined via a thorough sensitivity analysis, which incorporated evaluations of both horizontal pleiotropy and the impact of heterogeneity. The MR–Egger intercept test was used to detect horizontal pleiotropy ($P < 0.05$) when there were more than two instrumental variables. Heterogeneity was assessed by calculating the Cochran's Q statistic via both IVW and MR–Egger techniques [24].

2.4 Bayesian colocalization analysis

A Bayesian colocalization method was used to determine if traits are linked to the same pathogenic variants via *coloc* software's standard parameters. Colocalization between CRC cells and significant genes was assessed via the *coloc* application (<https://github.com/chr1swallace/coloc>) to understand the impact of signal loci. The analysis considered four scenarios: PP.H1: SNPs related to gene expression but not CRC susceptibility; PP.H2: related to CRC susceptibility without gene expression correlation; PP.H3: related to both but influenced by different SNPs; and PP.H4: associated with CRC susceptibility and gene expression, influenced by identical SNPs. The colocalization significance threshold is established at PP.H4 > 0.80, with genes that colocalize with CRC being viewed as potential therapeutic targets [25]. The colocalization results were visualized via "gwasglue" and "gassocplot" R scripts, v4.3.3.

2.5 Phenome-wide association analysis

To assess the potential pleiotropy and adverse effects of drug targets, an extensive phenotypic–genetic association analysis was conducted via AstraZeneca's PheWAS platform (<https://azphewas.com/>) [26]. This foundational research utilized information on approximately 15,500 dichotomous and 1,500 continuous phenotypic traits from a sample of approximately 450,000 individuals who underwent exome sequencing, according to the UK Biobank. The methods are detailed in the original publication [26].

2.6 PPI (protein–protein interaction) network construction

Examining PPI networks is crucial for understanding the complex interactions between proteins within cells. In this study, a PPI network was established via the STRING database with default parameters [27]. The PPI outcomes were subsequently visualized via Cytoscape software (version 3.1.9). The minimum required interaction score for the PPI analysis was 0.9.

2.7 Enrichment analysis

To elucidate the roles and biological importance of potential drug target genes, enrichment analyses were performed via the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. These studies utilized the R package clusterProfiler for comparing biological themes among gene clusters and Pathview, an R/Bioconductor tool for integrating and visualizing pathway-based data. GO annotations, which include biological process (BP), molecular function (MF), and cellular component (CC), systematically categorize gene functions. Moreover, KEGG provides information on the metabolic pathways related to these genes.

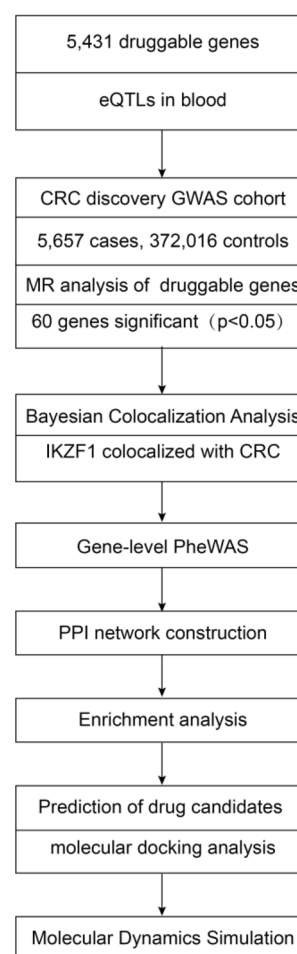
2.8 Candidate drug prediction and molecular docking

Assessing protein–pharmaceutical molecular interactions is key to determining the therapeutic potential of target genes. The Drug Signature Database (DSigDB) (<https://dsigdb.tanlab.org/>) is often consulted to map drugs and chemicals to genomic targets, aiding in the identification of drug candidates and their therapeutic potential.

In an effort to delve deeper into the interaction dynamics and druggability of target genes, molecular docking analyses have been conducted at the atomic level. These are crucial for evaluating the binding energies and interactions between drugs and targets, allowing for the selection of ligands with strong binding affinities for more efficient experimental validation and drug design. In this study, AutoDock Vina 1.1.2 was used for molecular docking simulations, and CB-Dock (<https://cadd.labshare.cn/cb-dock2/>) was employed to enhance the docking process by identifying binding sites and predicting conformations [17, 18]. The final step involved visualizing the blind docking results.

2.9 Molecular dynamics simulation

The Desmond module of the Schrodinger 2019 software was utilized for the present study, with the simulation of water molecules being conducted via the SPC water model, which was integrated with the OPLS2005 force field. To achieve a neutral system charge, a suitable quantity of chloride and sodium ions was introduced to counterbalance the charge and was randomly dispersed within the solvent environment. Following the assembly of the solvent-enriched system, energy minimization was performed via the standard procedure of the Desmond module, which incorporates parameters from the OPLS-2005 force field. Temperature and pressure control were achieved via the Nose–Hoover thermostat and isotropic scaling techniques, which were set to maintain a temperature of 300 Kelvin and a pressure equivalent to 1 atmosphere. Following this procedure, a series of NPT (nuclear parameter theory) simulations, each lasting 200 ns, were executed. The simulation data were recorded at intervals of 200 picoseconds to ensure the preservation of the trajectory for further analysis. The research flow of this study is shown in Fig. 1.

Fig. 1 Overview of the study design

3 Results

3.1 MR analysis identified 60 genes significantly associated with the likelihood of CRC onset

From the eQTLGen dataset, we identified cis-eQTLs associated with SNPs with F statistics greater than 10, serving as exposure data in MR analysis for strong correlations with exposure factors. The MR analysis included a filtering step, keeping genes with P values < 0.05 via the IVW/Wald ratio method and $P > 0.05$ in pleiotropy tests, indicating no significant horizontal pleiotropy (Tables S1, S2). The heterogeneity analysis (Table S3) revealed no heterogeneity among SNPs for any individual gene. The final screen identified 60 genes associated with CRC. Their Manhattan and forest plots are shown in Fig. 2.

3.2 Bayesian colocalization analysis

Studies indicate that significant MR effects may result from a genetic location with high linkage disequilibrium among SNPs, potentially leading to separate causal SNPs influencing SNP–exposure and SNP–outcome relationships and possibly causing false positives [28]. A collinearity assessment helps in determining if the exposure and the outcome are linked by a common causal SNP, especially when there is a pronounced correlation between SNPs and both the exposure and the outcome [25]. Empirical evidence suggests that proteins identified through MR and colocalization are more likely to be considered drug targets [29].

Fig. 2 **a** Mendelian randomization analysis revealed 60 genes associated with colorectal cancer risk. The y-axis represents the $-\log_{10}(p)$ of the association between gene expression and colorectal cancer risk. The x-axis represents the gene names and their corresponding chromosomal locations. Significant genes ($P < 0.05$) are labeled and highlighted in red. The red line represents the nominal significance threshold of 0.05. **b** Forest plots of the associations between 60 significant genes and colorectal cancer risk. The y-axis represents the odds ratio (95% CI) of the association between gene expression and colorectal cancer risk. The x-axis represents the gene names.

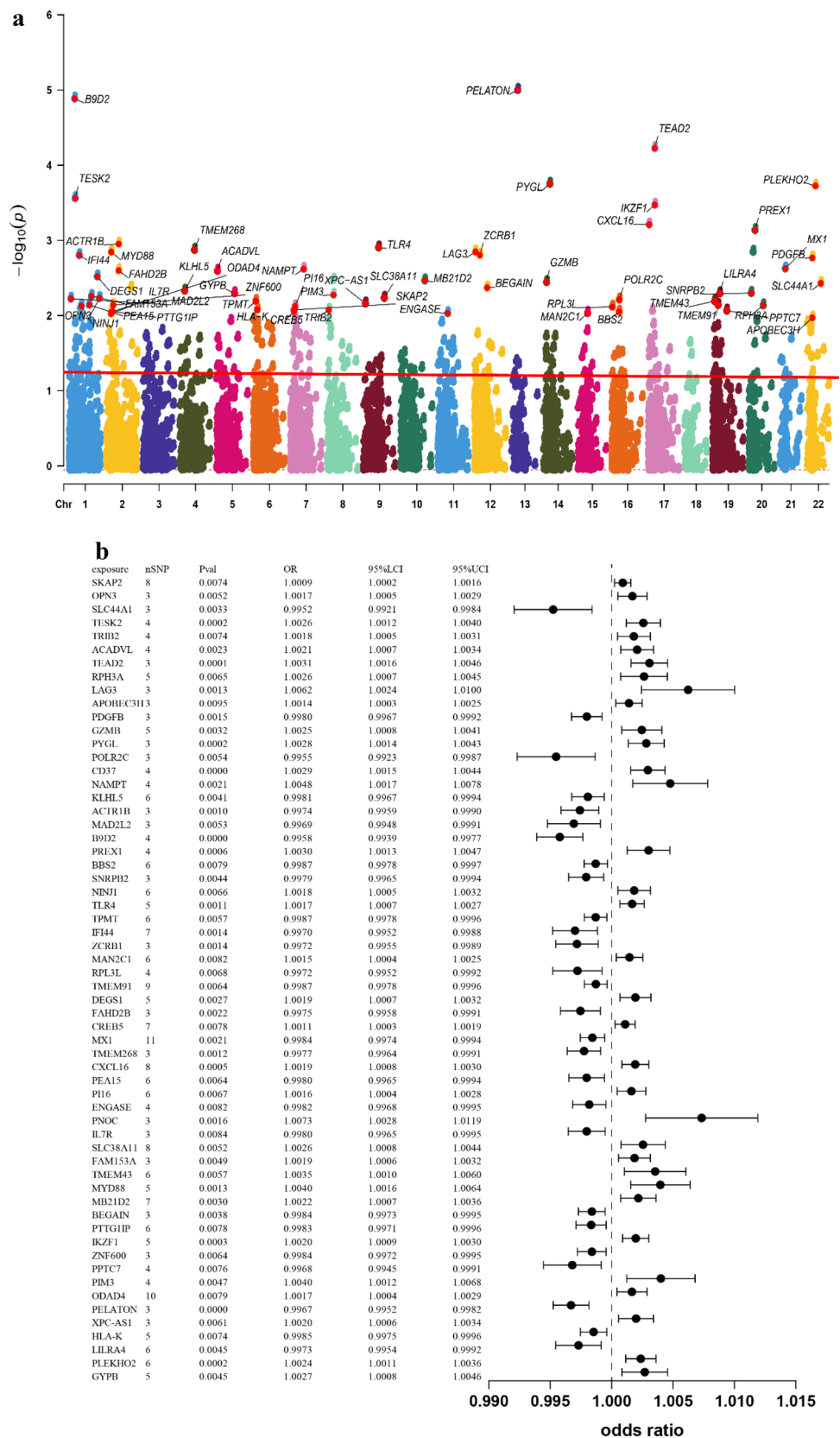
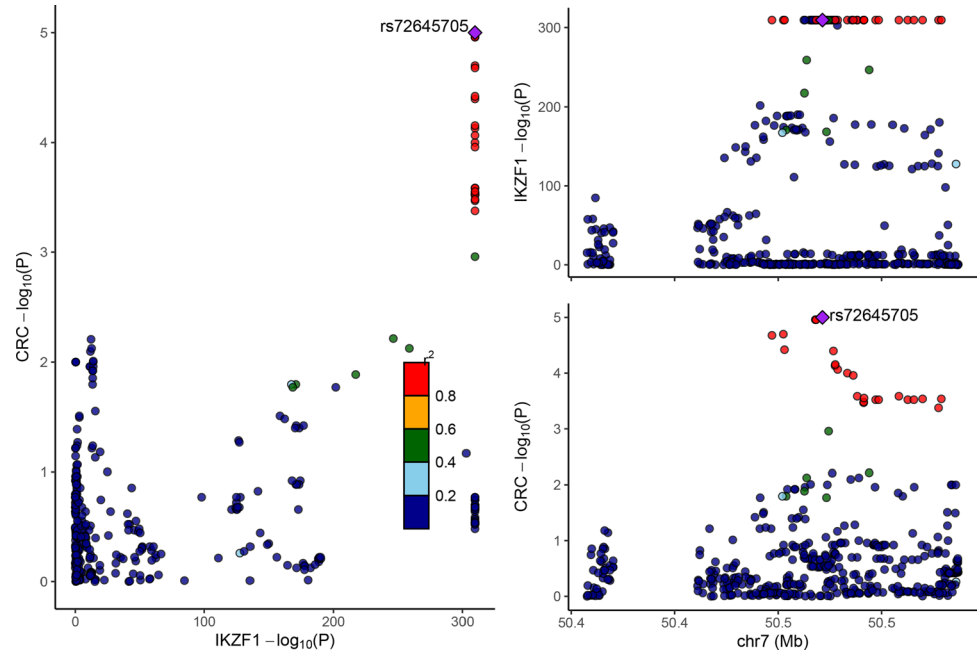


Table 1 Colocalization results of eQTLs for IKZF1 with CRC-associated SNPs

Gene	PP.H0	PP.H1	PP.H2	PP.H3	PP.H4
IKZF1	0.000	0.151	0.000	0.004	0.845

Fig. 3 Colocalization analysis revealed that the IKZF1 gene is significantly associated with colorectal cancer risk. The y-axis represents the posterior probability of the colocalization of IKZF1 gene expression and CRC risk. The x-axis represents the chromosomal location



In this research, Bayesian colocalization analysis was applied to 60 previously identified genes. Strikingly, the IKZF1 gene showed a pronounced colocalization signal (posterior probability of H4 > 0.8) linked to CRC (Table 1 and Fig. 3). As a result, IKZF1 has emerged as a promising candidate for drug target identification.

3.3 PheWAS

This study employed the AstraZeneca PheWAS Portal to conduct a comprehensive gene-level PheWAS. The results revealed no significant genomic correlations between IKZF1 and any other trait, adhering to a rigorous statistical threshold of $P < 5 \times 10^{-8}$ (Figures S1 and S2). This lack of correlation suggests that IKZF1 is less likely to have side effects or multiple significant effects on different biological pathways.

3.4 Construction of the PPI network

IKZF1 was input into the STRING (<https://cn.string-db.org/>) database to construct an interaction network, which was visualized with Cytoscape. PPI analysis revealed that IKZF1 interacts with 10 proteins, such as DDB1, CRBN, HDAC1, HDAC2, CHD4, SIN3A, GATA1, RUNX1, EBF1, and PAX5, resulting in an 11-node network. Visually, IKZF1’s ties with DDB1 and HDAC2 seemed weaker than other network interactions did (Fig. 4).

3.5 Enrichment analysis

GO enrichment analysis usually identifies gene–biological term links, whereas KEGG pathway enrichment analysis reveals gene participation in specific biological pathways [30].

GO enrichment analysis indicated that within the BP category, the pathways exhibiting the highest significance were related to the epigenetic modulation of gene expression. In the CC category, enrichment analysis revealed that IKZF1 is notably enriched within chromosomal regions and the histone deacetylase (HDAC) complex. Furthermore, in the MF category, IKZF1 is involved in functions that are intimately associated with transcriptional regulation (Fig. 5a).

Fig. 4 PPI network built with STRING. The blue circles represent IKZF1, and the green rectangles represent the associated genes

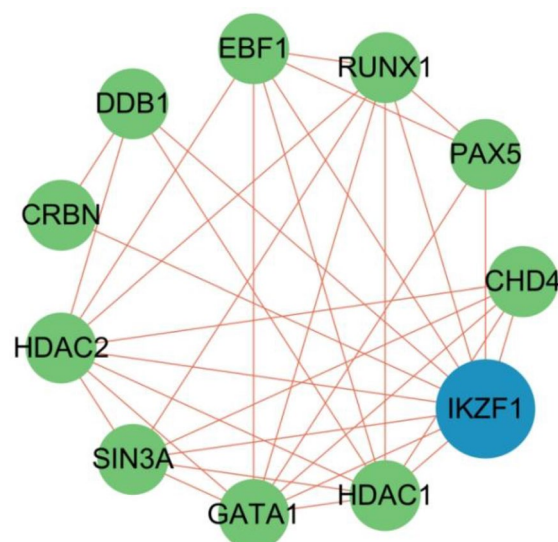


Figure 5b shows the KEGG pathway analysis highlighting transcriptional dysregulation in cancer. It includes various genetic changes in transcription factors, such as amplifications, deletions, rearrangements, and mutations, which can either increase or decrease their activity.

3.6 Prediction of drug candidates and molecular docking analysis

Using the DSigDB database, this study identified 50 potential drugs (Table S4). Molecular docking with AutoDock Vina v.1.2.2 was used to assess their interaction with IKZF1, and the binding sites and binding energy data were determined (Tables S5–S9). The docking results revealed that most complexes had strong binding affinities, averaging -7.028 kcal/mol, with those below -8 kcal/mol considered stable. Figure 6 depicts the top four binding complexes with IKZF1: imatinib mesylate, sanguinarine, lanatoside C, and lenalidomide, all with binding energies exceeding the average (Tables S6–S9). Notably, the binding energies for the digitalis derivatives (digitoxigenin and lanatoside C) and IKZF1 were less than -8 kcal/mol, indicating robust binding. Furthermore, the binding energies for the complexes of lanatoside C and imatinib mesylate with IKZF1 at all five binding sites were less than -8 kcal/mol (Tables S6, S8), indicating a high affinity for these interactions.

3.7 Molecular dynamics simulation

Molecular dynamics simulations demonstrated the stability of the IKZF1-ligand complexes in terms of conformation and interactions. The RMSD values of IKZF1 and the corresponding ligands varied acceptably within 3 \AA , and the range of fluctuations of the four IKZF1-ligand complexes was less than 3 \AA (Figure S3). During the simulation process, the four IKZF1-ligand complexes fluctuated greatly at the beginning, but all of them gradually stabilized at approximately 10 ns. Since the ligand has many rotatable bonds, the ligand fluctuates up and down from its binding site at the beginning and end of the simulation process, but the final RMSD of the four ligands is less than 3 \AA , which indicates that the small molecules are well bound to the protein pocket and do not undergo a large positional change.

The ligand's RMSF serves as a measure of its stability in conjunction with IKZF1 and the binding site. This metric quantifies the displacement of the ligand's atoms, while the highest RMSF value for IKZF1 pinpoints the protein's most dynamic segment over the course of the simulation. Throughout the simulation, the ligands in all four complexes exhibited fluctuations of less than 3 \AA , with Sanguinarine and Lenalidomide showing particularly minimal movement of less than 0.3 \AA (Figure S4). Additionally, the minor RMSF variations observed in IKZF1 across all four complexes suggest a high degree of stability (Figure S5).

The interactions between proteins and ligands can be categorized into four distinct categories: hydrogen bonds (depicted in green), hydrophobic interactions (shaded in pink), ionic interactions (colored red), and water bridge interactions (represented in blue). Analysis of the simulation data revealed that imatinib mesylate and lanatoside C predominantly exhibit water bridge interactions, alongside a balance of hydrophobic and hydrogen bonding, whereas

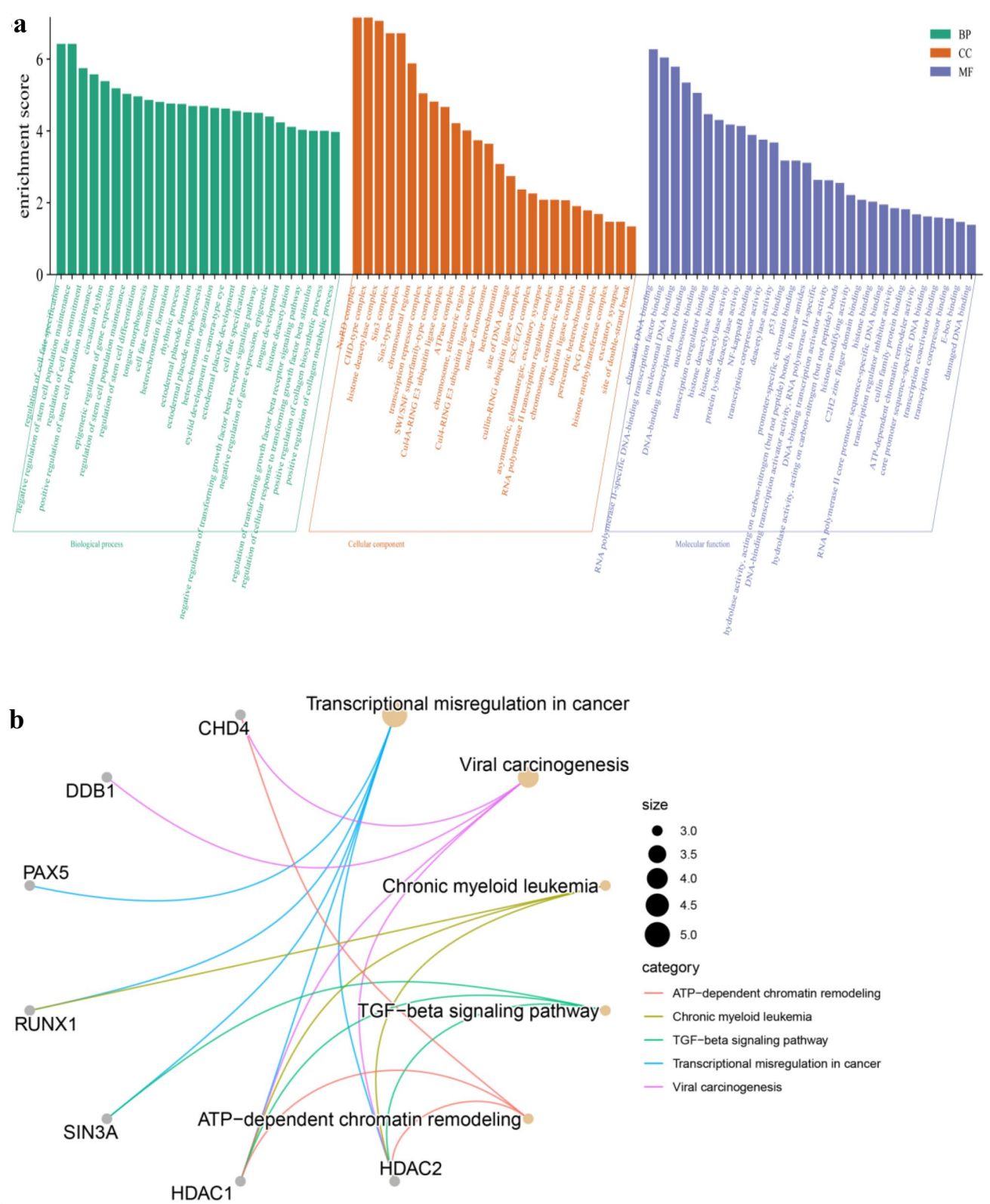


Fig. 5 **a** GO enrichment analysis reveals the role of the IKZF1 gene in epigenetic regulation and transcriptional control. The y-axis represents the gene ratio or enrichment score of the genes enriched in each GO term. The x-axis represents the GO terms categorized into biological process, molecular function, and cellular component. Different colors represent different GO term categories. **b** KEGG enrichment analysis revealed that the IKZF1 gene is involved in transcriptional dysregulation in cancer. The y-axis represents the gene ratio or enrichment score of the genes enriched in each KEGG pathway. The x-axis represents the KEGG pathways. Different colors represent different KEGG pathway categories

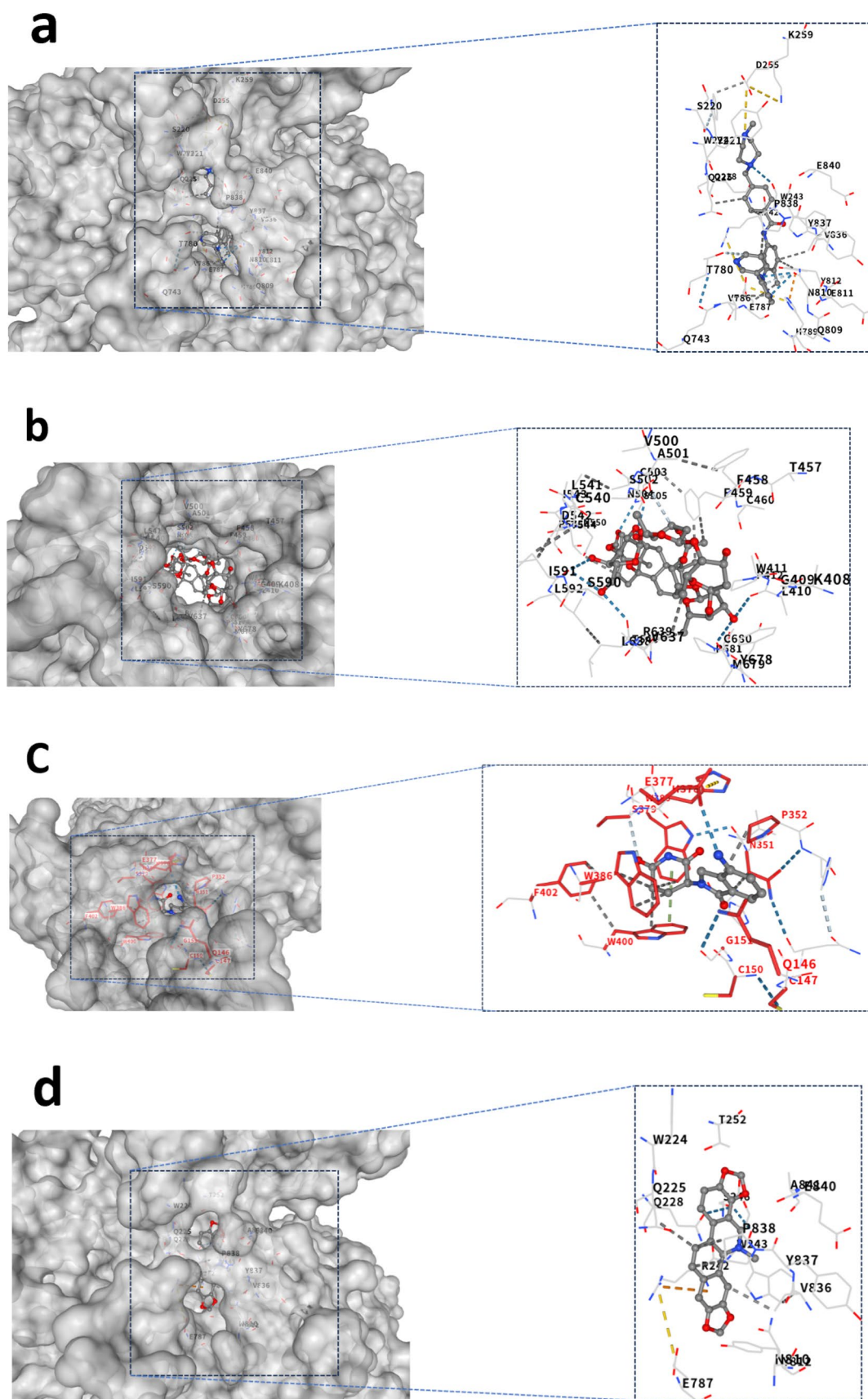


Fig. 6 **a** IKZF1 docking with imatinib mesylate. **b** IKZF1 docking with lanatoside C. **c** IKZF1 docking with lenalidomide. **d** IKZF1 docking with Sanguinarium.

sanguinarium and lenalidomide predominantly display hydrogen bonding, complemented by hydrophobic and water bridge interactions (Figure S6). Figure S7 delineates the cumulative number of ligand-specific interactions between IKZF1 and the ligand across the simulation trajectory, detailing the residues that engage with the ligand at each frame. Notably, certain residues form multiple specific interactions with the ligand.

The interactions between IKZF1 and its ligands are illustrated in Figure S8. Imatinib mesylate has an 87% hydrogen bonding interaction with the VAL356 residue of IKZF1; Lanatoside C exhibits an 82% hydrogen bonding interaction with the ALA501 residue of IKZF1; sanguinarium has a 52% hydrogen bonding interaction with the ASN810 residue of IKZF1; and lenalidomide has a 99% hydrogen bonding interaction with the TRP380 residue of IKZF1.

4 Discussion

The IKZF1 gene, located on chromosomal region 7p12.2, encompasses eight exons that are responsible for the production of the IKAROS protein [31, 32]. The IKZF1 gene family comprises five similar members, namely, Ikaros (IKZF1), Helios (IKZF2), Aiolos (IKZF3), Eos (IKZF4), and Pegasus (IKZF5). Molecularly, the IKAROS protein is distinguished by its six zinc finger domains. The key to its DNA-binding function is the N-terminal F1–F4 motif. Conversely, the C-terminal F5 and F6 regions are instrumental in enabling protein dimerization and interaction, which subsequently affects gene expression control. The phosphorylation of serine and threonine amino acids at both ends of a protein is a key regulatory mechanism that adjusts the DNA-binding activity of IKAROS during the developmental stages of lymphocytes. The precise coordination of these structural components and posttranslational modifications is essential for the correct execution of the role of IKAROS in the maturation process and lineage allocation of lymphocytes [34].

IKZF1 primarily plays a role in several key biological processes: it modulates gene expression and is involved in immune system regulation and the development of cancer [35]. These functions align with the outcomes of our enrichment analysis. With respect to gene expression control, research by Kenji Ichiyama and colleagues revealed that IKZF1 interacts with Foxp3 via the fifth exon, known as IKe5. They discovered that Treg cells lacking IKe5 exhibit increased expression of genes typically suppressed by Foxp3 upon T-cell receptor engagement, such as *Ifng*. This investigation revealed a significant correlation between IKZF1 and the functionality of Treg cells, as well as their role in autoimmune and anti-tumor responses [36]. IKZF1 functions as a modulator of genes targeted by the Notch signaling pathway, influencing chromatin reorganization. This influence can occur through direct interaction with these genes, leading to the establishment of heterochromatic regions, or by engaging other regulatory proteins, including HDAC1 [37]. Notably, our PPI analysis revealed a robust association between IKZF1 and HDAC1. Additionally, the expression levels of the IKZF1 gene are significantly associated with the maturation and functionality of various immune cell types. Studies by Thomas et al. and Bandyopadhyay et al. have elucidated the role of IKZF1 in modulating IL-2 levels in T cells. By suppressing IKZF1 in human CD3⁺ T cells, IKZF1 suppresses autocrine IL-2 production; conversely, IKZF1 suppression leads to an increase in IL-2 secretion [37]. Nevertheless, the expression of the IKZF1 gene is heterogeneous across various malignancies. Studies have revealed that IKZF1 tends to exhibit increased expression in acute myeloid leukemia and diffuse large B-cell lymphoma. In contrast, elevated levels of IKZF1 in cervical cancer, head and neck cancer, lung cancer, sarcoma, and cutaneous melanoma are associated with improved patient survival. Conversely, in low-grade glioma and uveal melanoma, elevated IKZF1 expression is linked to an unfavorable prognosis. In certain solid tumors, including hepatocellular carcinoma, the expression of IKZF1 is typically lower [33]. This finding implies a nuanced relationship between the expression patterns and roles of the IKZF1 gene across various cancer types.

Our enrichment analysis revealed that IKZF1 plays a pivotal role in the epigenetic control of gene expression. Research into the relationship between IKZF1 and colorectal cancer has focused on the use of IKZF1 methylation for the detection of initial cancer development and subsequent relapses. Studies have shown that there is a positive correlation between hypermethylation of IKZF1 and gene activity silencing, and CRC tumor tissues (but not adjacent nontumor tissues) present high methylation levels of IKZF1 at all stages [38]. The methylation status of the IKZF1 gene has also been associated with the risk of recurrence of CRC, and the predictive value of the risk of recurrence can be improved when this status is combined with clinicopathological factors [39]. A comparative study demonstrated that the methylated BCAT1/IKZF1 gene test outperforms the conventional carcinoembryonic antigen (CEA) test in identifying CRC recurrence, and the IKZF1 methylation status is proposed as a standalone prognostic marker for CRC [40]. While most of these studies have focused on IKZF1 methylation and CRC, Bhanu Sharma et al. reported an elevated level of IKZF1 gene expression (rs6964823) in colorectal tissue samples from individuals residing in Jammu and Kashmir. Importantly, correlation is not causal, and many genetic and environmental

factors could influence the associations found in a specific population. However, our study with a wide European cohort partially supports a causal link between IKZF1 and CRC, which is in line with the findings of Bhanu Sharma et al. Together, these findings suggest that IKZF1 is a major CRC therapeutic target with significant potential for developing targeted gene therapies.

A strong interaction affinity between imatinib mesylate and IKZF1 was observed from the molecular docking results and molecular dynamics simulations. As a type of selective tyrosine kinase inhibitor, imatinib mesylate exerts its effects by targeting specific tyrosine kinases. This therapeutic effect is achieved primarily by inhibiting the tyrosine kinase activity that is linked to the BCR–ABL fusion proteins [41]. Sunitinib has been shown to suppress the growth of colon fibroblasts both in cellular cultures and within living organisms by specifically disrupting the PDGFR signaling cascade. These findings indicate that sunitinib could serve as a therapeutic agent for CRC by impeding the activity of tumor-associated stromal fibroblasts, thus offering a novel strategy for the multifaceted management of CRC [42]. Although the present study focused on sunitinib rather than imatinib, it underscores the potential of tyrosine kinase inhibitors such as sunitinib in CRC therapy. Despite molecular docking studies suggesting stable binding, current research on the specific interaction mechanism of imatinib mesylate with IKZF1 is rare. However, the findings from this study may facilitate future investigations into the therapeutic potential of imatinib mesylate for treating CRC.

Through our drug prediction and molecular docking analyses, we found that digitalis analogs such as digitoxigenin and lanatoside C had binding energies < -8 kcal/mol with IKZF1, indicating strong interactions. Molecular dynamics simulations also confirmed this. This study provides initial evidence for the stable binding of digitalis to IKZF1, suggesting its potential as a therapeutic agent for CRC. Emerging research suggests a potential function for digitalis analogs in gastrointestinal tumor therapy. Research indicates that cardiac glycosides (CGs) can increase the sensitivity of liver cancer cells to the apoptotic ligand TRAIL. This is achieved through ROS production, activation of the p38MAPK pathway, inhibition of survival proteins, interruption of autophagy, and a protective response [43]. Dhanasekhar Reddy et al. reported that lanatoside C has anticancer activity in human cancer cells (breast, lung, and liver) by affecting multiple signaling pathways, including the MAPK, Wnt/ β -catenin, JAK-STAT, and PI3K/AKT/mTOR pathways [44]. In addition, lanatoside C-treated cancer cells exhibited increased apoptosis and reduced cell viability under ER stress conditions, findings that support further investigation of CGs as potential antitumor agents for the treatment of pancreatic and other cancers that depend on GRP78 for growth and survival [45]. The dysregulation of IKZF1, an essential transcription factor, in various cancers is strongly correlated with tumor progression and patient prognosis. Therapeutically targeting IKZF1 might present new strategies for CRC treatment. Despite the scant literature on the direct link between digitalis, IKZF1, and CRC, our research provides a foundation for additional exploration in this field.

Our results confirmed a strong and stable interaction between IKZF1 and sanguinarine, which is consistent with the findings of prior studies. Sanguinarine reportedly induces apoptosis in HT-29 human CRC cells by altering the Bax/Bcl-2 ratio and initiating a caspase-9-dependent pathway. These findings suggest that sanguinarine can promote apoptosis in CRC cells, which may provide substantial therapeutic advantages for CRC treatment [46]. However, few studies have elucidated the interaction mechanisms between sanguinarine and IKZF1.

Research has indicated that imidazoline derivatives (ImiDs), including thalidomide, lenalidomide, and pomalidomide, exert an inhibitory effect on the proliferation of multiple myeloma (MM) cells. This process is accomplished through the enhancement of ubiquitination and subsequent breakdown of vital transcription factors, specifically IKZF1 and IKZF3, in MM cells. This process involves the binding of ImiDs to CRBN, which leads to the degradation of IKZF1. This degradation, in turn, diminishes the expression levels of IRF4 and MYC, thereby impeding their contribution to the proliferation of MM cells. This action results in an enhanced immune response, particularly in T cells, which is crucial for the body's defense mechanisms against MM [47]. While the research in question did not directly focus on CRC, the findings of this study could shed light on the potential application of lenalidomide in CRC therapy, considering its recognized function as an immunomodulatory agent across a spectrum of malignancies.

This study has several advantages. It leverages MR to discover CRC therapeutic targets through a genetic lens, tapping into the most expansive publicly available GWAS dataset on CRC risk to date. The enrichment analysis revealed the functional attributes and regulatory interplay of the genes. The conclusive drug prediction underscored the therapeutic efficacy of these genes, and the elevated binding affinity observed via molecular docking and molecular dynamics simulations confirmed their substantial potential as drug targets. This study provides a thorough evaluation, spanning from target discovery to binding characteristics, and identifies IKZF1 as a promising drug target with compelling evidence for CRC.

5 Limitations of the study

The study is not without its constraints. While instrumental for causal inference, MR is predicated on assumptions of minimal drug dosage and a direct correlation between exposure and outcome, which might not align with the high-dose, short-term assessments typical of clinical trials. Consequently, MR might not precisely mirror the therapeutic effects observed in clinical practice. Moreover, the reliance of this study on plasma eQTL data for MR analysis could present obstacles in pinpointing optimal therapeutic targets across various tissues, each with unique genetic control mechanisms. The study's scope is also somewhat confined by its predominantly European ancestry cohort, necessitating further investigation to ascertain the applicability of these findings across diverse ethnicities. MR is not impervious to biases stemming from unmeasured confounders or pleiotropic effects, which could skew the study's conclusions. It is imperative to acknowledge these limitations and their potential influence on the study's outcomes. Furthermore, the study's emphasis on cis-eQTLs and their link to CRC might have disregarded other genetic and environmental elements that contribute to the intricacy of the disease. While enrichment analysis is beneficial, it is predicated on predefined gene sets or pathways, which may not encompass all biological processes or interactions. The lack of significant enrichment should not be construed as an absence of biological significance, and caution is warranted in interpreting these results. Finally, the fidelity of molecular docking is contingent upon the accuracy of the protein and ligand structures. While this method can identify prospective drug targets, it does not ascertain their clinical efficacy, which requires further empirical validation and clinical trials.

6 Conclusion

In summary, our investigation utilized MR to identify IKZF1 as a prospective therapeutic target in colorectal cancer, which was supported by gene colocalization analyses. The gene's link to immune modulation and epigenetic mechanisms underscores its potential for therapy. Moreover, molecular docking studies and molecular dynamics simulations reinforced the suitability of IKZF1 for drug targeting, underscoring its role in enhancing precision CRC therapies. This research notably enriches the therapeutic options for colorectal cancer, encouraging further exploration and clinical trials of treatments aimed at IKZF1.

Acknowledgements Not applicable.

Author contributions Junzuo Lin's contributions: Study design, data collection, data analysis, interpretation of results, and writing of the first draft of the paper. Qian Xu's Contribution: Research design, execution of experiments, and preliminary analysis of the data. Zhengfei Zhao's Contribution: Design of research concepts, interpretation of data, and writing and revising draft papers.

Funding The authors state that no monetary assistance was obtained for the study, the writing of the manuscript, or the publication of this article.

Data availability Data is provided within the manuscript or supplementary information files. The datasets analyzed during the current study are available in the following repositories: eQTL data were obtained from the eQTLGen Consortium (<https://eqtlgen.org/>); CRC data were obtained from the UK Biobank and downloaded from the GWAS Catalog (<https://gwas.mrcieu.ac.uk/datasets/ieu-b-4965/>).

Declarations

Ethics approval and consent to participate All studies involving humans were evaluated and endorsed by the National Center for Health Statistics, and the individuals involved provided informed consent in writing to participate in the research.

Consent for publication The data used in this study are from a public database and do not contain any personally identifiable information. Therefore, consent for publication was not needed.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If

material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209–49. <https://doi.org/10.3322/caac.21660>.
2. Hossain MS, Karuniawati H, Jairoun AA, Urbi Z, Ooi J, John A, Lim YC, Kibria KMK, Mohiuddin AKM, Ming LC, et al. Colorectal cancer: a review of carcinogenesis, global epidemiology, current challenges, risk factors, preventive and treatment strategies. *Cancers (Basel)*. 2022;14(7):1732. <https://doi.org/10.3390/cancers14071732>.
3. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, Floratos A, Sham PC, Li MJ, Wang J, et al. The support of human genetic evidence for approved drug indications. *Nat Genet*. 2015;47(8):856–60. <https://doi.org/10.1038/ng.3314>.
4. Hingorani AD, Kuan V, Finan C, Kruger FA, Gaulton A, Chopade S, Sofat R, MacAllister RJ, Overington JP, Hemingway H, et al. Improving the odds of drug development success through human genomics: modelling study. *Sci Rep*. 2019;9(1):18911. <https://doi.org/10.1038/s41598-019-54849-w>.
5. King EA, Davis JW, Degner JF. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genet*. 2019;15(12): e1008489. <https://doi.org/10.1371/journal.pgen.1008489>.
6. Schmidt AF, Finan C, Gordillo-Marañón M, Asselbergs FW, Freitag DF, Patel RS, Tyl B, Chopade S, Faraway R, Zwierzyńska M, et al. Genetic drug target validation using Mendelian randomisation. *Nat Commun*. 2020;11(1):3255. <https://doi.org/10.1038/s41467-020-16969-0>.
7. Finan C, Gaulton A, Kruger FA, Lumbers RT, Shah T, Engmann J, Galver L, Kelley R, Karlsson A, Santos R, et al. The druggable genome and support for target identification and validation in drug development. *Sci Transl Med*. 2017. <https://doi.org/10.1126/scitranslmed.aag1166>.
8. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, Li Y, Kurreeman FA, Zhernakova A, Hinks A, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet*. 2010;42(6):508–14. <https://doi.org/10.1038/ng.582>.
9. Okada Y, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, Kawaguchi T, Stahl EA, Kurreeman FA, Nishida N, et al. Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. *Nat Genet*. 2012;44(5):511–6. <https://doi.org/10.1038/ng.2231>.
10. Ishigaki K, Sakaue S, Terao C, Luo Y, Sonehara K, Yamaguchi K, Amariuta T, Too CL, Laufer VA, Scott IC, et al. Multi-ancestry genome-wide association analyses identify novel genetic mechanisms in rheumatoid arthritis. *Nat Genet*. 2022;54(11):1640–51. <https://doi.org/10.1038/s41588-022-01213-w>.
11. Katan MB. Apolipoprotein E isoforms, serum cholesterol, and cancer. *Lancet*. 1986;1(8479):507–8. [https://doi.org/10.1016/s0140-6736\(86\)92972-7](https://doi.org/10.1016/s0140-6736(86)92972-7).
12. Smith GD, Ebrahim S. “Mendelian randomization”: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003;32(1):1–22. <https://doi.org/10.1093/ije/dyg070>.
13. Holmes MV, Ala-Korpela M, Smith GD. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. *Nat Rev Cardiol*. 2017;14(10):577–90. <https://doi.org/10.1038/nrcardio.2017.78>.
14. Hingorani A, Humphries S. Nature's randomised trials. *Lancet*. 2005;366(9501):1906–8. [https://doi.org/10.1016/s0140-6736\(05\)67767-7](https://doi.org/10.1016/s0140-6736(05)67767-7).
15. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, Montgomery GW, Goddard ME, Wray NR, Visscher PM, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016;48(5):481–7. <https://doi.org/10.1038/ng.3538>.
16. Storm CS, Kia DA, Almramhi MM, Bandres-Ciga S, Finan C, Hingorani AD, Wood NW. Finding genetically-supported drug targets for Parkinson's disease using Mendelian randomization of the druggable genome. *Nat Commun*. 2021;12(1):7342. <https://doi.org/10.1038/s41467-021-26280-1>.
17. Liu Y, Yang X, Gan J, Chen S, Xiao ZX, Cao Y. CB-Dock2: improved protein-ligand blind docking by integrating cavity detection, docking and homologous template fitting. *Nucleic Acids Res*. 2022;50(W1):W159–w164. <https://doi.org/10.1093/nar/gkac394>.
18. Yang X, Liu Y, Gan J, Xiao ZX, Cao Y. FitDock: protein-ligand docking by template fitting. *Brief Bioinform*. 2022. <https://doi.org/10.1093/bib/bbac087>.
19. Vösa U, Claringbould A, Westra HJ, Bonder MJ, Deelen P, Zeng B, Kirsten H, Saha A, Kreuzhuber R, Yazar S, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet*. 2021;53(9):1300–10. <https://doi.org/10.1038/s41588-021-00913-z>.
20. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, et al. The MR-Base platform supports systematic causal inference across the human phenotype. *Elife*. 2018. <https://doi.org/10.7554/eLife.34408>.
21. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ*. 2018;362: k601. <https://doi.org/10.1136/bmj.k601>.
22. Burgess S, Thompson SG. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol*. 2011;40(3):755–64. <https://doi.org/10.1093/ije/dyr036>.
23. Burgess S, Davey Smith G, Davies NM, Dudbridge F, Gill D, Glymour MM, Hartwig FP, Kutalik Z, Holmes MV, Minelli C, et al. Guidelines for performing Mendelian randomization investigations: update for summer 2023. *Wellcome Open Res*. 2019;4:186. <https://doi.org/10.12688/wellcomeopenres.15555.3>.
24. Greco MF, Minelli C, Sheehan NA, Thompson JR. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Stat Med*. 2015;34(21):2926–40. <https://doi.org/10.1002/sim.6522>.

25. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 2014;10(5): e1004383. <https://doi.org/10.1371/journal.pgen.1004383>.
26. Wang Q, Dhindsa RS, Carss K, Harper AR, Nag A, Tachmazidou I, Vitsios D, Deevi SVV, Mackay A, Muthas D, et al. Rare variant contribution to human disease in 281,104 UK Biobank exomes. *Nature.* 2021;597(7877):527–32. <https://doi.org/10.1038/s41586-021-03855-y>.
27. Otasek D, Morris JH, Bouças J, Pico AR, Demchak B. Cytoscape Automation: empowering workflow-based network analysis. *Genome Biol.* 2019;20(1):185. <https://doi.org/10.1186/s13059-019-1758-4>.
28. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum Mol Genet.* 2018;27(R2):R195–r208. <https://doi.org/10.1093/hmg/ddy163>.
29. Zheng J, Haberland V, Baird D, Walker V, Haycock PC, Hurler MR, Gutteridge A, Erola P, Liu Y, Luo S, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat Genet.* 2020;52(10):1122–31. <https://doi.org/10.1038/s41588-020-0682-6>.
30. Chen L, Zhang YH, Lu G, Huang T, Cai YD. Analysis of cancer-related lncRNAs using gene ontology and KEGG pathways. *Artif Intell Med.* 2017;76:27–36. <https://doi.org/10.1016/j.artmed.2017.02.001>.
31. Georgopoulos K, Bigby M, Wang JH, Molnar A, Wu P, Winandy S, Sharpe A. The Ikaros gene is required for the development of all lymphoid lineages. *Cell.* 1994;79(1):143–56. [https://doi.org/10.1016/0092-8674\(94\)90407-3](https://doi.org/10.1016/0092-8674(94)90407-3).
32. Molnár A, Georgopoulos K. The Ikaros gene encodes a family of functionally diverse zinc finger DNA-binding proteins. *Mol Cell Biol.* 1994;14(12):8292–303. <https://doi.org/10.1128/mcb.14.12.8292-8303.1994>.
33. Yang LK, Lin CX, Li SH, Liang JJ, Xiao LL, Xie GH, Liu HW, Liao X. Novel IKZF3 transcriptomic signature correlates with positive outcomes of skin cutaneous melanoma: a pan-cancer analysis. *Front Genet.* 2022;13:1036402. <https://doi.org/10.3389/fgene.2022.1036402>.
34. Paolino J, Tsai HK, Harris MH, Pikman Y. IKZF1 alterations and therapeutic targeting in B-cell acute lymphoblastic leukemia. *Biomedicines.* 2024. <https://doi.org/10.3390/biomedicines12010089>.
35. Feng L, Zhang H, Liu T. Multifaceted roles of IKZF1 gene, perspectives from bench to bedside. *Front Oncol.* 2024;14:1383419. <https://doi.org/10.3389/fonc.2024.1383419>.
36. Ichiyama K, Long J, Kobayashi Y, Horita Y, Kinoshita T, Nakamura Y, Kominami C, Georgopoulos K, Sakaguchi S. Transcription factor Ikzf1 associates with Foxp3 to repress gene expression in Treg cells and limit autoimmunity and anti-tumor immunity. *Immunity.* 2024. <https://doi.org/10.1016/j.immuni.2024.07.010>.
37. Awwad MHS, Kriegsmann K, Plaumann J, Benn M, Hillengass J, Raab MS, Bertsch U, Munder M, Weisel K, Salwender HJ, et al. The prognostic and predictive value of IKZF1 and IKZF3 expression in T-cells in patients with multiple myeloma. *Oncoimmunology.* 2018;7(10): e1486356. <https://doi.org/10.1080/2162402x.2018.1486356>.
38. Jedi M, Young GP, Pedersen SK, Symonds EL. Methylation and gene expression of BCAT1 and IKZF1 in colorectal cancer tissues. *Clin Med Insights Oncol.* 2018;12:1179554918775064. <https://doi.org/10.1177/1179554918775064>.
39. Al Naji H, Winter JM, Pedersen SK, Roy A, Byrne SE, Young GP, Symonds EL. Evaluating the role of methylated circulating tumor DNA in combination with pathological prognostic factors for predicting recurrence of colorectal cancer. *Biomark Insights.* 2024;19:11772719241232870. <https://doi.org/10.1177/11772719241232870>.
40. Bhattarai A, Shah S, Khadka P, Serhan HA, Sah R, Sah S. Methylated BCAT1/IKZF1 DNA: a breakthrough in colorectal cancer diagnosis? *Ann Med Surg (Lond).* 2024;86(1):11–2. <https://doi.org/10.1097/ms9.0000000000001495>.
41. Iqbal N, Iqbal N. Imatinib: a breakthrough of targeted therapy in cancer. *Chemother Res Pract.* 2014;2014: 357027. <https://doi.org/10.1155/2014/357027>.
42. Wang ZH, Li Q, Ruan SQ, Xiao Q, Liu Y, Hu YT, Hu LF, Chen HY, Zheng S, Zhang SZ, et al. Sunitinib mesylate inhibits proliferation of human colonic stromal fibroblasts in vitro and in vivo. *J Zhejiang Univ Sci B.* 2014;15(8):701–12. <https://doi.org/10.1631/jzus.B1300306>.
43. Rasheduzzaman M, Yin H, Park SY. Cardiac glycoside sensitized hepatocellular carcinoma cells to TRAIL via ROS generation, p38MAPK, mitochondrial transition, and autophagy mediation. *Mol Carcinog.* 2019;58(11):2040–51. <https://doi.org/10.1002/mc.23096>.
44. Reddy D, Kumavath R, Ghosh P, Barh D. Lanatoside C induces G2/M cell cycle arrest and suppresses cancer cell growth by attenuating MAPK, Wnt, JAK-STAT, and PI3K/AKT/mTOR signaling pathways. *Biomolecules.* 2019;9(12):792. <https://doi.org/10.3390/biom9120792>.
45. Ha DP, Tsai YL, Lee AS. Suppression of ER-stress induction of GRP78 as an anti-neoplastic mechanism of the cardiac glycoside Lanatoside C in pancreatic cancer: lanatoside C suppresses GRP78 stress induction. *Neoplasia.* 2021;23(12):1213–26. <https://doi.org/10.1016/j.neo.2021.10.004>.
46. Lee JS, Jung WK, Jeong MH, Yoon TR, Kim HK. Sanguinarine induces apoptosis of HT-29 human colon cancer cells via the regulation of Bax/Bcl-2 ratio and caspase-9-dependent pathway. *Int J Toxicol.* 2012;31(1):70–7. <https://doi.org/10.1177/1091581811423845>.
47. Wang S, Jin FY. Advances on immunomodulatory drugs against multiple myeloma. *Zhonghua Xue Ye Xue Za Zhi.* 2016;37(3):262–4. <https://doi.org/10.3760/cma.j.issn.0253-2727.2016.03.021>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.